Elsevier $\LaTeX$ template[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], *Global Customer Service*[b,*]

[a] *1600 John F Kennedy Boulevard, Philadelphia*
[b] *360 Park Avenue South, New York*

Abstract

This template helps you to create a properly formatted $\LaTeX$ manuscript.

Keywords:

1. Introduction

Suppose $X_1, \ldots, X_n$ are i.i.d. from p-dimensional normal distribution $N_p(\mu_X, \Sigma_X)$. Denote $X = (X_1, \ldots, X_n)$. In this paper, it is assumed that $n < p$, that is, high dimension setting is considered. Consider a linear regression model

$$y = \beta_0 \mathbf{1}_n + X^T \beta + \epsilon, \tag{1}$$

where $\mathbf{1}_n$ is n dimensional vector with all elements equal to 1 and ϵ has distribution $N(0, \sigma^2 I_n)$.

Let $\Sigma_X = P \Lambda P^T$ be the spectral decomposition of Σ_X, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and P is an orthogonal matrix. In PCA context, it is assumed that Σ_X is spiked, that is $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_p$ for some $r > 0$ (See [1]). Denote by P_1 the first r column of P and P_2 the last $p - r$ column of P. The aim of PCA is to estimate P_1. In this paper, we

[☆] Fully documented templates are available in the elsarticle package on CTAN.

[*] Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)

[1] Since 1880.

allow Σ_X to be either spiked or non-spiked. Non-spike means that there's no principal component ($r = 0$). That is, $\lambda_1 = \cdots = \lambda_p$. Spike means that there's r principal components for $r > 0$. In either case, let $\lambda = \lambda_{r+1} = \cdots = \lambda_p$.

If Σ_X is indeed spiked,

$$y = \beta_0 \mathbf{1}_n + X^T P_1 P_1^T \beta + X^T P_2 P_2^T \beta + \epsilon, \tag{2}$$

where $X^T P_1$ and $X^T P_2$ are independent. PCR try to do regression between y and $X^T P_1$. Since P_1 is not observed, it is substituted by an estimator $\tilde{P}_1$. Traditionally, PCR is a technique for analyzing multiple regression data that suffers from multicollinearity. Recently, PCR is a practical method to deal with high dimensional regression. If $p < n$, the full multicollinearity phenomenon shows up even if predictors are independent. It calls for a test procedure to justify the appropriateness of PCR. To be precise, we consider testing the hypotheses

$$H : \Sigma \text{ is non-spiked or } \Sigma \text{ is spiked and } P_1^T \beta = 0 \tag{3}$$

versus

$$K : \Sigma \text{ is spiked and } P_1^T \beta \neq 0. \tag{4}$$

If P_1 is observed, then the problem is reduced to testing an ordinary regression model. However, it's not the case. In fact, the classical F-test statistic for the regression between y and $X^T \tilde{P}_1$ may not be a good choice for at least three reasons:

1. From equation

   $$y = \beta_0 \mathbf{1}_n + X^T \tilde{P}_1 \tilde{P}_1^T \beta + X^T (I_p - \tilde{P}_1 \tilde{P}_1^T) \beta + \epsilon, \tag{5}$$

   we can see that the F-test suffers from Endogeneity.

2. The estimator of P_1 may not be consistent in high dimension. Moreover, Σ_X may not be spiked and, as a result, there's no principal component.

3. Even if there's additional information or data to estimate P_1, we will never know weather we estimate P_1 well enough such that the F-test is valid.

[2] proposed a generalized likelihood ratio test (GLRT) for testing high dimensional mean values. Roughly speaking, GLRT projects data to lower dimension by a direction a such that likelihood ratio is maximized. GLRT is likelihood based, it can be regarded as a generalization of classical LRT in high dimension setting.

In this paper we apply the GLRT method to the problem of testing the significance of PCR.

2. New Test

It can be seen that $(X_1^T, y_1)^T, \ldots, (X_n^T, y_n)^T$ are i.i.d. from $N_{p+1}(\mu, \Sigma)$, where $\mu = (\mu_X^T, \beta_0)^T$ and

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_X \beta \\ \beta^T \Sigma_X & \beta^T \Sigma_X \beta + \sigma^2 \end{pmatrix}. \tag{6}$$

Denote $\Theta : (\mu, \Sigma)$. Define the hypothesis H_a by

$$H_a : \mathrm{Cov}(a^T X_i, y_i) = 0, \tag{7}$$

where $a \in \mathbb{R}^p$ and $a^T a = 1$. Let

$$S = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} (X_i - \bar{X})(X_i - \bar{X})^T & (X_i - \bar{X})(y_i - \bar{y})^T \\ (y_i - \bar{y})(X_i - \bar{X})^T & (y_i - \bar{y})(y_i - \bar{y})^T \end{pmatrix} = \begin{pmatrix} S_{XX} & S_{Xy} \\ S_{yX} & S_{yy} \end{pmatrix}, \tag{8}$$

and

$$S_a = \begin{pmatrix} a^T & 0 \\ 0 & 1 \end{pmatrix} S \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_a = \begin{pmatrix} a^T & 0 \\ 0 & 1 \end{pmatrix} \Sigma \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}. \tag{9}$$

The likelihood function of $(a^T X_i, y_i)$, $i = 1, \ldots, n$, is

$$L_a(\theta; X, Y) = (2\pi)^{-n} |\Sigma_a|^{-n/2} \exp\left(-\frac{1}{2} \mathrm{tr} \Sigma_a^{-1} S_a\right). \tag{10}$$

Then the maximum likelihood is

$$L(a) = \sup_{\theta \in \Theta} L_a(\theta; X, Y) = (2\pi)^{-n} |S_a|^{-n/2} e^{-n}. \tag{11}$$

If $|S_a| = 0$, then 11 is interpreted as $+\infty$. Similarly, the maximum likelihood under H_a is

$$L(a) = \sup_{\theta \in H} L_a(\theta; X, Y) = (2\pi)^{-n} |a^T S_{XX} a S_{yy}|^{-n/2} e^{-n}. \tag{12}$$

In [2], GLRT is defined as

$$\min_{L(a)=+\infty} L_H(a) \quad s.t. \quad a^T a = 1. \tag{13}$$

The idea of GLRT is to find a such that $L(a) = +\infty$ and $L_H(a) < +\infty$ as small as possible such that the discrepancy between the likelihood values $L(a)$ and $L_H(a)$ is maximized. We call the direction a^* obtained by (13) the GLRT direction.

From the expression of $L(a)$ and $L_H(a)$, a^* is equal to

$$a^* = \text{argmax}_{a^T a=1} a^T S_{XX} a \quad s.t. \quad |S_a| = 0. \tag{14}$$

Such a direction a^* can be expected to make $|\Sigma_a|$ small and $a^T \Sigma_{XX} a$ large. That is, the variance of $a^T X_i$ is large and $a^T X_i$ and y_i are highly correlated. If X_i has certain principal components which are correlated to y_i, the direction a^* is expected to be close to corresponding principal directions.

Next we solve the optimization problem (14). Let $Q_n = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$. Denote by $Q_n = WW^T$ the rank decomposition of Q_n, where W_n is an $n \times (n-1)$ matrix with $W^T W = I_{n-1}$. Then $|S_a| = 0$ is equivalent to $a^T XQX^T a y^T Qy = (a^T XQy)^2$ and is equivalent to $W^T X^T a = W^T yk$ for some $k \in \mathbb{R}$. It follows that

$$a = XW(W^T X^T XW)^{-1} W^T yk + (I - XW(W^T X^T XW)^{-1} W^T X^T)a. \tag{15}$$

Since $a^T a = 1$,

$$k^2 y^T W(W^T X^T XW)^{-1} W^T y + a^T (I - XW(W^T X^T XW)^{-1} W^T X^T)a = 1. \tag{16}$$

Note that

$$L_H(a) \propto (a^T XQX^T a y^T Qy)^{-n/2} = \left(k^2 (y^T Q_n y)^2\right)^{-n/2}. \tag{17}$$

To make $L_H(a)$ minimized, we should maximize k^2. So the second term of 16 should be 0. That is

$$a = XW(W^TX^TXW)^{-1}W^Tyk \tag{18}$$

Hence

$$k^2 = \frac{1}{y^TW(W^TX^TXW)^{-1}W^Ty}, \tag{19}$$

and

$$L_H(a) \propto (a^TXQX^Tay^TQy)^{-n/2} = \Big(\frac{(y^TQ_ny)^2}{y^TW(W^TX^TXW)^{-1}W^Ty}\Big)^{-n/2}. \tag{20}$$

After homogenization, we define

$$T = \frac{y^TQ_ny}{y^TW(W^TX^TXW)^{-1}W^Ty}.$$

If T is large, we reject H.

3. Main Results

Let $\tilde{y} = W^Ty$, $\tilde{X} = XW$, $\tilde{\epsilon} = W^T\epsilon$. Then the columns of $\tilde{X}$ are i.i.d. distributed as $N(0, \Sigma_X)$, $\tilde{\epsilon} \sim N(0, \sigma^2 I_{n-1})$ and $\tilde{y} = \tilde{X}^T\beta + \tilde{\epsilon}$. The test statistic can be written as

$$T = \frac{\tilde{y}^T\tilde{y}}{\tilde{y}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{y}}.$$

We make the following assumption.

Assumption 1. *Assume model* (1) *holds with the columns of X i.i.d. distributed as $N(\mu_X, \Sigma_X)$, $\epsilon \sim N(0, \sigma^2 I_n)$ and σ^2 is fixed as $n, p \to \infty$.*

Assumption 2. *Assume the eigenvalues of Σ_X satisfy $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_p = \lambda$, where $r \geq 0$ and $\lambda > 0$ are fixed as $n, p \to \infty$, $\lambda_1 \asymp \lambda_r$ and $p^{1/2}/\lambda_r \to 0$. We say there's no principal component if $r = 0$, that is $\lambda_1 = \cdots = \lambda_p$.*

The null hypotheses H is the union of two disjoint hypothesis $H = \cup_{i=1}^2 H_i$, where H_1: There's no principal component; and H_2: There's r principal components with $r > 0$ and $P_1^T\beta = 0$. Under H_1 we have the following theorem

Theorem 1. *Suppose Assumptions 1 and 2 hold. Assume $p/n \to \infty$ and H_1 is true. Then*

$$T/(\lambda p) \xrightarrow{P} 1. \tag{21}$$

If $\|\beta\|^2 = o(\frac{n}{p})$ or $\|\beta\|^{-2} = o(\frac{n}{p})$, then for $\alpha \in (0,1)$ we have

$$\Pr\Big(\frac{T-\lambda p}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) \leq \alpha. \tag{22}$$

Remark 1. It can be seen from our proof that if $\|\beta\|^2 = o(\frac{n}{p})$ and $\|\beta\|^{-2} = o(\frac{n}{p})$ are both fail, then the asymptotic property of T is sophisticated.

Under H_2 we have a similar theorem with one more condition $p = o(n^2)$.

Theorem 2. *Suppose Assumptions 1 and 2 hold. Assume $p/n \to \infty$, $p/n^2 \to 0$ and H_2 is true. Then*

$$T/(\lambda p) \xrightarrow{P} 1. \tag{23}$$

If $\|P_2^T\beta\|^2 = o(\frac{n}{p})$ or $\|P_2^T\beta\|^{-2} = o(\frac{n}{p})$, then for $\alpha \in (0,1)$ we have

$$\Pr\Big(\frac{T-\lambda p}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) \leq \alpha. \tag{24}$$

Under K, to simplify the proof, we assume β is generated from a normal prior distribution before data are generated. Denote by $\Phi(\cdot)$ the CDF of normal distribution. We have the following theorem:

Theorem 3. *Suppose Assumptions 1 and 2 hold. Assume $p/n \to \infty$ and K is true. Assume β has prior distribution $N(0, \sigma_\beta^2 I_p)$. Asuume that*

$$\frac{np^2 + p^{5/2} + \lambda_1 p^{3/2}}{(p+\lambda_1)^2}\sigma_\beta^2 \to \infty, \tag{25}$$

then

$$\Pr\Big(\frac{T-p\lambda}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) = E\Phi\Big(-\Phi^{-1}(1-\alpha) + \frac{\sum_{i=1}^r \lambda_i\chi_i^2}{\lambda\sqrt{2p}}\Big) + o(1). \tag{26}$$

Remark 2. If $\lambda_1/\sqrt{p} \to \infty$, then

$$\Pr\Big(\frac{T-p\lambda}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) \to 1. \tag{27}$$

Since λ is unkown, an estimator should be substituted. A natural estimator is

$$\frac{1}{(p-r)(n-1)}\sum_{i=r+1}^{n-1}\lambda_i(\tilde{X}^T\tilde{X}). \tag{28}$$

However, r is unknown, which itself need to be estimated consistently. If $r>0$, it can be well estimated (See [3]). However, if $r=0$, which may occur in our problem, the method in [3] fails. Nevertheless, it's easy to find an estimator which is not less than r. In following theorem, we will assume k is a fixed number such that $k\geq r$.

Theorem 4. *Suppose Assumptions 1 and 2 hold. Assume $p/n\to\infty$ and $k\geq r$ is a fixed integer. Let*

$$\frac{1}{(p-k)(n-1)}\sum_{i=k+1}^{n-1}\lambda_i(\tilde{X}^T\tilde{X}). \tag{29}$$

Then

$$\hat{\lambda}=\lambda+O_P(\frac{1}{n}). \tag{30}$$

Furthermore, if we add condition $p=o(n^2)$ to Theorem 1, add condition $(p+\lambda_1)/(n\sqrt{p})\to 0$ to Theorem 3, then the conclusion of Theorem 1, 2 and 3 holds with λ substituted by $\hat{\lambda}$.

Remark 3. It is not hard to generalize our results for a random positive integer k such that $\Pr(k\geq r)\to 1$ and $k\leq M$ for some $M>0$.

Remark 4. In practice, r is often small. Hence it's often the case that we could choose a known upper bound for r.

By our theoretic results, we reject the hypotheses when

$$\frac{T-p\lambda}{\lambda\sqrt{2p}}\geq\Phi^{-1}(1-\alpha). \tag{31}$$

Under the condition of Theorem 1 and Theorem 2, the test level can be guaranteed. The test power is given by Theorem 3.

4. Appendix

For random variables ξ and η, we write $\xi \sim \eta$ when ξ and η have the same distribution. For two sequences of positive random variables ξ_n and η_n, we write $\xi_n \asymp \eta_n$ if $\Pr(c\eta_n \le \xi_n \le C\eta_n) \to 1$ for some positive c and C.

$$\begin{aligned} T &= \frac{\beta^T \tilde{X} \tilde{X}^T \beta + 2\beta^T \tilde{X} \tilde{\epsilon} + \tilde{\epsilon}^T \tilde{\epsilon}}{\beta^T \tilde{X} {(\tilde{X}^T \tilde{X})}^{-1} \tilde{X}^T \beta + 2\beta^T \tilde{X} {(\tilde{X}^T \tilde{X})}^{-1} \tilde{\epsilon} + \tilde{\epsilon}^T {(\tilde{X}^T \tilde{X})}^{-1} \tilde{\epsilon}} \\ &= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3}. \end{aligned} \tag{32}$$

4.1. Lemma

Lemma 1. *Suppose A is an $n \times n$ full rank symmetric matrix. And let*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{33}$$

where A_{11} is a real number, A_{12} is a $1 \times (n-1)$ matrix, A_{21} is a $(n-1) \times 1$ matrix and A_{22} is a $(n-1) \times (n-1)$ matrix. Denote $A_{11\cdot 2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$. Then we have

$$\left(A^{-1}\right)_{11} = A_{11\cdot 2}^{-1} \tag{34}$$

Lemma 2. *Let H and P be two symmetric matrices and $M = H + P$. If $j + k - n \ge i \ge r + s - 1$, we have*

$$\lambda_j(H) + \lambda_k(P) \le \lambda_i(M) \le \lambda_r(H) + \lambda_s(P). \tag{35}$$

Lemma 2 is known as the Weyl's inequality.

Lemma 3. *Suppose $B = \frac{1}{q} V V^T$ where V is an $p \times q$ random matrix composed of i.i.d. random variables with zero mean, unit variance and finite fourth moment. As $q \to \infty$ and $p/q \to c \in [0, +\infty)$, the largest and smallest nonzero eigenvalues of B converge almost surely to ${(1+\sqrt{c})}^2$ and ${(1-\sqrt{c})}^2$, respectively.*

Lemma 3 is known as the Bai-Yin's law [4].

Lemma 4. *Let $Z_1, \dots, Z_{n+1}$ i.i.d. distributed as $N(0, I_p)$. $\Lambda = diag(\lambda_1, \dots, \lambda_p)$, where $\lambda_1 \geq \cdots \lambda_r$ and $\lambda_{r+1} = \cdots = \lambda_p = \lambda$. $\limsup_{n\to\infty} \lambda_1/\lambda_r < \infty$, $\lambda_1/\sqrt{p} \to \infty$. Suppose $p = o(n^2)$. Denote $Z = (Z_1, \dots, Z_n)$. Let $\hat{V}$ be the first r eigenvectors of $\Lambda^{1/2} Z Z^T \Lambda^{1/2}$, $V = (e_1, \dots, e_r)$. Then*

$$Z_{n+1}^T \Lambda^{1/2} (V V^T - \hat{V}\hat{V}^T) \Lambda^{1/2} Z_{n+1} = o(\sqrt{p}) \tag{36}$$

Lemma 4 is from Wang Rui's paper.

Lemma 5. *Suppose $F_n(\cdot)$ and $F(\cdot)$ are distribution functions and $F_n \xrightarrow{L} F$, then*

$$\sup_x |F_n(x) - F(x)| \to 0. \tag{37}$$

See Exercise 3.2.9 of [5].

Lemma 6. *Suppose Z is an $p \times n$ ($p \geq n$) random matrix with all elements i.i.d. distributed as $N(0,1)$. Denote by $Z = U\Lambda V^T$ the singular value decomposition (SVD) of Z, where U is a $p \times n$ orthogonal matrix, Λ is an $n \times n$ diagonal matrix and V is an $n \times n$ orthogonal matrix. Then U, Λ and V are independent. (See, e.g., [6])*

Lemma 7. *Let A be an $n \times n$ symmetric positive semi-definite matrix with rank r. Denote by $A = P\Lambda P^T$ the spectral decomposition of A, where P is an $n \times r$ orthogonal matrix and $\Lambda = diag(\lambda_1, \dots, \lambda_r)$ is an $r \times r$ diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r > 0$. Then we have*

$$(A + I_n)^{-1} \geq I_n - PP^T \tag{38}$$

Proof. Let $\tilde{P}$ be an $n \times n$ orthogonal matrix such that P is the first r columns of $\tilde{P}$. And let $\tilde{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ be an $n \times n$ matrix. Then $P\Lambda P^T = \tilde{P}\tilde{\Lambda}\tilde{P}^T$, and

$$\begin{aligned}
(A + I_n)^{-1} &= \tilde{P}(\tilde{\Lambda} + I_n)^{-1} \tilde{P}^T \\
&= \tilde{P}\mathrm{diag}((\lambda_1 + 1)^{-1}, \dots, (\lambda_r + 1)^{-1}, 1, \dots, 1)\tilde{P}^T \\
&\geq \tilde{P}\mathrm{diag}(0, \dots, 0, 1, \dots, 1)\tilde{P}^T \\
&= I_n - PP^T
\end{aligned} \tag{39}$$

$\square$

4.2. circumstance 1

4.2.1. First randomization of β

Independent of data, generate a random p dimensional orthonormal matrix O with Haar invariant distribution. And

$$T = \frac{(O\beta)^T O\tilde{X}(O\tilde{X})^T O\beta + 2(O\beta)^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}(O\tilde{X})^T\beta + 2(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon}} \tag{40}$$

Note that conditioning on O, $O\tilde{X}$ is a random matrix with each entry independently distributed as $N(0,\lambda)$. Hence O is independent of $O\tilde{X}$. Observe also that $O\beta/\|\beta\|$ is uniformly distributed on the unit ball. We can without loss of generality and assume that $\beta/\|\beta\|$ is uniformly distributed on the surface unit ball in (32).

4.2.2. Second randomization of β

Independent of data, generate $R > 0$ with R^2 distributed as χ^2_p. Then $\xi = R\beta/\|\beta\|$ distributed as $N_p(0, I_p)$. Note that conditioning on $\tilde{X}$, $\eta = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\xi$ is distributed as $N_{n-1}(0, I_{n-1})$. Hence η is independent of $\tilde{X}$.

Then

$$\begin{aligned} T &= \frac{(\|\beta\|/R)^2\xi^T\tilde{X}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\ &= \frac{(\|\beta\|/R)^2\eta^T\tilde{X}^T\tilde{X}\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\eta^T\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\ &= \frac{A_1 + A_2 + A_3}{B_1 + B_2 + B_3} \end{aligned} \tag{41}$$

4.2.3. Step 3: CLT

Similar to the derivation of the distribution of Hotelling's T^2 statistic.

Now we deal with

$$\frac{A_3}{B_3} = \frac{\tilde{\epsilon}^T \tilde{\epsilon}}{\tilde{\epsilon}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{\epsilon}} \tag{42}$$

Let O be an $(n-1) \times (n-1)$ orthogonal matrix satisfies

$$O\tilde{\epsilon} = \begin{pmatrix} \|\tilde{\epsilon}\| \\ 0 \\ \cdots \\ 0 \end{pmatrix}.$$

Then

$$\frac{A_3}{B_3} = \frac{(O\tilde{\epsilon})^T O\tilde{\epsilon}}{(O\tilde{\epsilon})^T \left((\tilde{X}O^T)^T \tilde{X}O^T\right)^{-1} O\tilde{\epsilon}}. \tag{43}$$

It can be seen that $\tilde{X}O^T$ has the same distribution as $\tilde{X}$ and is independent of O. We have

$$\frac{A_3}{B_3} \sim \frac{1}{((\tilde{X}^T \tilde{X})^{-1})_{11}}. \tag{44}$$

Apply Lemma 1, we have

$$\frac{A_3}{B_3} \sim (\tilde{X}^T \tilde{X})_{11\cdot 2}. \tag{45}$$

Since $\tilde{X}^T \tilde{X} \sim \text{Wishart}_{n-1}(\lambda I_{n-1}, p)$, $(\tilde{X}^T \tilde{X})_{11\cdot 2} \sim \lambda \chi^2_{p-n+2}$. Hence $A_3/B_3 \asymp p$ and

$$\frac{A_3/B_3 - \lambda(p-n+2)}{\lambda\sqrt{2(p-n+2)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{46}$$

by CLT.

Similar technique can deal with A_1/B_1:

$$\frac{A_1}{B_1} = \frac{\eta^T \tilde{X}^T \tilde{X} \eta}{\eta^T \eta} \sim (\tilde{X}^T \tilde{X})_{11} \sim \lambda \chi^2_p. \tag{47}$$

Hence $A_1/B_1 \asymp p$ and

$$\frac{A_1/B_1 - \lambda p}{\lambda\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1), \tag{48}$$

by CLT.

4.2.4. step 4

It's obvious that $A_3 \asymp n$ and $B_1 \asymp \frac{n}{p}\|\beta\|^2$. We already have $A_1/B_1 \asymp p$ and $A_3/B_3 \asymp p$. It follows that $A_1 \asymp n\|\beta\|^2$ and $B_3 \asymp n/p$. And

$$\begin{aligned} A_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T(\tilde{X}^T\tilde{X})\eta} \\ &= O_P(\|\beta\|/\sqrt{p})O_P(\sqrt{np}) \\ &= O_P(\sqrt{n}\|\beta\|), \end{aligned} \tag{49}$$

$$\begin{aligned} B_2 &= O_P(\|\beta\|/\sqrt{p})\eta^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} \\ &= O_P(\|\beta\|/\sqrt{p})\sqrt{\eta^T(\tilde{X}^T\tilde{X})^{-1}\eta} \\ &= O_P(\|\beta\|/\sqrt{p})O_P(\sqrt{n/p}) \\ &= O_P(\frac{\sqrt{n}}{p}\|\beta\|). \end{aligned} \tag{50}$$

Note that

$$A_2 = O_P(\frac{1}{\sqrt{n}})n\|\beta\| = O_P(\frac{1}{\sqrt{n}})\sqrt{A_1}\sqrt{A_3} \le O_P(\frac{1}{\sqrt{n}})(A_1 + A_3) \tag{51}$$

Similarly we have $B_2 = O_P(\frac{1}{\sqrt{n}})(B_1 + B_3)$. It follows that

$$T = \frac{A_1 + A_3}{B_1 + B_3}(1 + O_P(\frac{1}{\sqrt{n}})). \tag{52}$$

For every $\epsilon > 0$, we have

$$\begin{aligned} &\Pr\Big(\frac{A_1 + A_3}{B_1 + B_3} \ge (\lambda p)(1+\epsilon)\Big) \\ =&\Pr(A_1 + A_3 \ge (B_1 + B_3)(\lambda p)(1+\epsilon)) \\ \le&\Pr(A_1 \ge B_1(\lambda p)(1+\epsilon)) + \Pr(A_3 \ge B_3(\lambda p)(1+\epsilon)). \end{aligned} \tag{53}$$

But

$$\frac{A_1}{\lambda p B_1} \xrightarrow{P} 1 \quad \text{and} \quad \frac{A_3}{\lambda p B_3} \xrightarrow{P} 1. \tag{54}$$

It follows that (53) tends to 0. Similarly,

$$\Pr\Big(\frac{A_1 + A_3}{B_1 + B_3} \le (\lambda p)(1-\epsilon)\Big) \to 0. \tag{55}$$

We have proved

$$\frac{A_1 + A_3}{\lambda p(B_1 + B_3)} \xrightarrow{P} 1. \tag{56}$$

Together with (52), it follows that $T \xrightarrow{P} 1$.

4.2.5. Step 5

By Cauchy inequility, $\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}\tilde{\epsilon}^T\tilde{X}^T\tilde{X}\tilde{\epsilon} \geq (\tilde{\epsilon}^T\tilde{\epsilon})^2$. Denote $B_4 = \tilde{\epsilon}^T\tilde{X}^T\tilde{X}\tilde{\epsilon}$. Using similar technique as before, we have $B_4 \asymp np$ and $(B_4/A_3 - \lambda p)/(\lambda\sqrt{2p}) \xrightarrow{\mathcal{L}} N(0,1)$. Together with (52) and (56), we have

$$\begin{aligned}
T &= \frac{A_1 + A_3}{B_1 + B_3} \frac{1 + \frac{A_2}{A_1 + A_3}}{1 + \frac{B_2}{B_1 + B_3}} \\
&= \frac{A_1 + A_3}{B_1 + B_3}(1 + \frac{A_2}{A_1 + A_3})(1 - \frac{B_2}{B_1 + B_3}(1 + o_P(1))) \\
&= \frac{A_1 + A_3}{B_1 + B_3}(1 + (\frac{|A_2|}{A_1 + A_3} + \frac{|B_2|}{B_1 + B_3})(1 + o_P(1))) \\
&= \frac{A_1 + A_3}{B_1 + B_3} + O_P(p)(\frac{|A_2|}{A_1 + A_3} + \frac{|B_2|}{B_1 + B_3}) \\
&= \frac{A_1 + A_3}{B_1 + B_3} + O_P(p)(\frac{\sqrt{n}\|\beta\|}{n\|\beta\|^2 + n} + \frac{\frac{\sqrt{n}}{p}\|\beta\|}{\frac{n}{p}\|\beta\|^2 + \frac{n}{p}}) \\
&= \frac{A_1 + A_3}{B_1 + B_3} + O_P(\frac{p\sqrt{n}\|\beta\|}{n\|\beta\|^2 + n}) \\
&\leq \frac{A_1 + A_3}{B_1 + A_3^2/B_4} + O_P(\frac{p\sqrt{n}\|\beta\|}{n\|\beta\|^2 + n}).
\end{aligned} \tag{57}$$

We deal with the two terms seperately.

$$\frac{\frac{A_1 + A_3}{B_1 + A_3^2/B_4} - \lambda p}{\lambda\sqrt{2p}} = c\frac{A_1/B_1 - \lambda p}{\lambda\sqrt{2p}} + (1 - c)\frac{B_4/A_3 - \lambda p}{\lambda\sqrt{2p}}, \tag{58}$$

where

$$c = \frac{B_1}{B_1 + A_3^2/B_4} \asymp \frac{\frac{n}{p}\|\beta\|^2}{\frac{n}{p}\|\beta\|^2 + \frac{n}{p}} = \frac{\|\beta\|^2}{\|\beta\|^2 + 1}. \tag{59}$$

Hence by Slutsky's theorem, we have

$$\frac{\frac{A_1 + A_3}{B_1 + A_3^2/B_4} - \lambda p}{\lambda\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1), \tag{60}$$

if $\|\beta\| \to 0$ or $\|\beta\| \to \infty$.

To control the second term of (57), we further require

$$\frac{\sqrt{np}\|\beta\|}{n\|\beta\|^2+n} \to 0. \tag{61}$$

Equivalently, if $\|\beta\| \to 0$, we require $\|\beta\| = o(\frac{\sqrt{n}}{\sqrt{p}})$; if $\|\beta\| \to \infty$, we require $\|\beta\|^{-1} = o(\frac{\sqrt{n}}{\sqrt{p}})$.

If these conditions are satiesfied, we have

$$\Pr\Big(\frac{T-\lambda p}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) \leq \alpha \tag{62}$$

4.3. circumstance 2

Assumption 3. $P_1^T\beta = 0$.

$$\begin{aligned}
T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{\beta^TP_2P_2^T\tilde{X}\tilde{X}^TP_2P_2^T\beta + 2\beta^TP_2P_2^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^TP_2P_2^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^TP_2P_2^T\beta + 2\beta^TP_2P_2^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1+A_2+A_3}{B_1+B_2+B_3}
\end{aligned} \tag{63}$$

4.3.1. Step 1

Like before, we have $A_3/B_3 \sim (\tilde{X}^T\tilde{X})_{11\cdot 2}$. Denote by $\Lambda = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)$. Let $Z = (Z_1,\ldots,Z_p)$ be a $n-1\times p$ matrix with all elements independently distributed as $N(0,1)$. Let $Z_{(1)}$ and $Z_{(2)}$ be the first 1 row and last $n-2$ rows of Z, that is

$$Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}.$$

Then

$$\begin{aligned}
\tilde{X}^T\tilde{X} &\sim Z\Lambda Z^T \\
&= \begin{pmatrix} Z_{(1)}\Lambda Z_{(1)}^T & Z_{(1)}\Lambda Z_{(2)}^T \\ Z_{(2)}\Lambda Z_{(1)}^T & Z_{(2)}\Lambda Z_{(2)}^T \end{pmatrix}.
\end{aligned} \tag{64}$$

Hence

$$
\begin{aligned}
A_3/B_3 &\sim Z_{(1)}\Lambda Z_{(1)}^T - Z_{(1)}\Lambda Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1}Z_{(2)}\Lambda Z_{(1)}^T \\
&= Z_{(1)}\Lambda^{1/2}\big(I_p - \Lambda^{1/2}Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1}Z_{(2)}\Lambda^{1/2}\big)\Lambda^{1/2}Z_{(1)}^T \qquad (65)\\
&\le Z_{(1)}\Lambda^{1/2}\big(I_p - \hat{V}\hat{V}^T\big)\Lambda^{1/2}Z_{(1)}^T,
\end{aligned}
$$

where $\hat{V}$ is the first r eigenvectors of $\Lambda^{1/2}Z_{(2)}^T Z_{(2)}\Lambda^{1/2}$. From PCA theory (see [1]), $\hat{V}\hat{V}^T$ is a good estimator of population principal space VV^T even in high dimensional setting. Here $V = (e_1, \ldots, e_r)$, where e_i is the vector with all elements equal to 0 but the ith equal to 1. Note that we have required $p = o(n^2)$. Then by lemma 4,

$$Z_{(1)}\Lambda^{1/2}\big(VV^T - \hat{V}\hat{V}^T\big)\Lambda^{1/2}Z_{(1)}^T = o(\sqrt{p}). \qquad (66)$$

Note that

$$Z_{(1)}\Lambda^{1/2}\big(I - VV^T\big)\Lambda^{1/2}Z_{(1)}^T \sim \lambda\chi^2_{p-r} \qquad (67)$$

Hence $A_3/B_3 \le \lambda\chi^2_{p-r} + o(\sqrt{p})$.

On the other hand, the non-zero eigenvalues of $\Lambda^{1/2}\big(I_p - \Lambda^{1/2}Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1}Z_{(2)}\Lambda^{1/2}\big)\Lambda^{1/2}$ is no less than that of $\lambda(I_p - \Lambda^{1/2}Z_{(2)}^T\big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1}Z_{(2)}\Lambda^{1/2})$. Hence $A_3/B_3 \ge \lambda\chi^2_{p-n+2}$.

It follows that $A_3/B_3 \asymp p$ if $p/n \to \infty$.

4.3.2. Step 2: B_1 and B_2

Note that $P_2^T\tilde{X}$ is an $(p-r)\times(n-1)$ matrix with all elements independently distributed as $N(0,\lambda)$. Similar to non-spiked circumstance, we have $A_1 \asymp n\|P_2^T\beta\|^2$, $A_2 = O_P(\sqrt{n}\|P_2^T\beta\|)$, $A_3 \asymp n$ and $B_3 \asymp n/p$.

Next we deal with B_1. Let $P_2^T\tilde{X} = U_2D_2V_2^T$ be the SVD of $P_2^T\tilde{X}$, where U_2 is a $(p-r)\times(n-1)$ orthonormal matrix, D_2 is a $(n-1)\times(n-1)$ diagonal matrix and V_2 is a $(n-1)\times(n-1)$ orthonormal matrix. Without loss of generality, we can assume $P_2^T\beta/\|P_2^T\beta\|$ is uniformly distributed on the surface of unit ball.

B_1 has the following upper bound:

$$\begin{aligned} B_1 &\leq \beta^T P_2 P_2^T \tilde{X} (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta \\ &= \beta^T P_2 U_2 U_2^T P_2^T \beta \end{aligned} \tag{68}$$

Independent of $P_2^T \beta / \|P_2^T \beta\|$ and U_2, we generate $R \sim \chi^2_{p-r}$. Then we have

$$\sqrt{R} \frac{P_2^T \beta}{\|P_2^T \beta\|} \sim N_{p-r}(0, I_{p-r}). \tag{69}$$

Hence

$$\begin{aligned} &\beta^T P_2 U_2 U_2^T P_2^T \beta \\ &= \frac{\sqrt{R} \beta^T P_2}{\|P_2 \beta^T\|} U_2 U_2^T \frac{\sqrt{R} P_2^T \beta}{\|P_2^T \beta\|} \frac{1}{R} \|P_2^T \beta\|^2 \\ &\asymp \frac{n-1}{p-r} \|P_2^T \beta\|^2. \end{aligned} \tag{70}$$

To get the lower bound, note that

$$\begin{aligned} B_1 &= \beta^T P_2 P_2^T \tilde{X} (\tilde{X}^T P_1 P_1^T \tilde{X} + \tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta \\ &= \beta^T P_2 U_2 D_2 V_2^T (\tilde{X}^T P_1 P_1^T \tilde{X} + V_2 D_2^2 V_2^T)^{-1} V_2 D_2 U_2^T P_2^T \beta \\ &= \beta^T P_2 U_2 (D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1} + I_{n-1})^{-1} U_2^T P_2^T \beta. \end{aligned} \tag{71}$$

Here U_2 is independent of $(V_2, D_2, P_1^T \tilde{X})$. By lemma 7

$$(D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1} + I_{n-1})^{-1} \geq I_{n-1} - U^* U^{*T} \tag{72}$$

where U^* is the first r eigenvectors of $D_2^{-1} V_2^T \tilde{X}^T P_1 P_1^T \tilde{X} V_2 D_2^{-1}$ and is independent of U_2. Since U_2 has Haar distribution, we have

$$\begin{aligned} B_1 &\geq \beta^T P_2 U_2 (I_{n-1} - U^* U^{*T}) U_2^T P_2^T \beta \\ &= \beta^T P_2 U_2 U_2^T P_2^T \beta - \beta^T P_2 U_2 U^* U^{*T} U_2^T P_2^T \beta. \end{aligned} \tag{73}$$

The difference of upper bound and lower bound is

$$\beta^T P_2 U_2 U^* U^{*T} U_2^T P_2^T \beta \asymp \frac{r}{p-r} \|P_2^T \beta\|^2. \tag{74}$$

Hence

$$\begin{aligned} B_1 &= \beta^T P_2 P_2^T \tilde{X} (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta + O_p(\frac{r}{p-r} \|P_2^T \beta\|^2) \\ &= \beta^T P_2 P_2^T \tilde{X} (\tilde{X}^T P_2 P_2^T \tilde{X})^{-1} \tilde{X}^T P_2 P_2^T \beta (1 + O_P(1/n)). \end{aligned} \tag{75}$$

So that $B_1 \asymp \frac{n}{p}\|P_2^T\beta\|^2$.

For B_2 we have

$$\begin{aligned} B_2 &= O_P(1)\sqrt{\beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T\tilde{X})^{-2}\tilde{X}^T P_2 P_2^T \beta} \\ &\leq \lambda_{\min}(\tilde{X}^T\tilde{X})^{-1/2} O_P(1)\sqrt{\beta^T P_2 P_2^T \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T P_2 P_2^T \beta}. \end{aligned} \tag{76}$$

But $\lambda_{\min}(\tilde{X}^T\tilde{X}) \geq \lambda_{\min}(\tilde{X}^T P_2 P_2^T \tilde{X}) \asymp p - r$ by Lemma 3. Hence $B_2 = O_P(\frac{\sqrt{n}}{p}\|P_2^T\beta\|)$.

Hence the similar law of large number and CLT holds.

4.3.3. Step 3

Use similar technique as before, we have

$$\frac{A_1}{B_1} \sim \frac{\chi_p^2}{1 + O_P(1/n)} = \lambda\chi_p^2(1 + O_P(1/n)). \tag{77}$$

It follows by large number that

$$\frac{A_1/B_1}{\lambda p} \xrightarrow{P} 1. \tag{78}$$

And if $p = o(n^2)$, we have

$$\frac{A_1/B_1 - \lambda p}{\lambda\sqrt{2p}} \sim \frac{\chi_p^2(1 + O_P(1/n)) - p}{\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{79}$$

Recall that if $p = o(n^2)$, we have $A_3/B_3 \geq \lambda\chi_{p-n+2}^2$ and $A_3/B_3 \leq \lambda\chi_{p-r}^2 + o(\sqrt{p}) \leq \lambda\chi_p^2 + o(\sqrt{p})$. Then

$$\frac{A_3/B_3}{\lambda p} \xrightarrow{P} 1, \tag{80}$$

and

$$\frac{A_3/B_3 - \lambda p}{\lambda\sqrt{2p}} \leq \frac{\chi_p^2 + o(\sqrt{p}) - p}{\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{81}$$

From (78) and (80) we can deduce $T/(\lambda p) \xrightarrow{P} 1$ by similar argument as before.

Similar to (57), we have

$$T \leq \frac{A_1 + A_3}{B_1 + A_3} + O_P(\frac{p\sqrt{n}\|P_2^T\beta\|}{n\|P_2^T\beta\|^2 + n}). \tag{82}$$

For the first term, we have

$$\frac{\frac{A_1+A_3}{B_1+B_3}-\lambda p}{\lambda\sqrt{2p}} = c\frac{\frac{A_1}{B_1}-\lambda p}{\lambda\sqrt{2p}} + (1-c)\frac{\frac{A_3}{B_3}-\lambda p}{\lambda\sqrt{2p}}, \tag{83}$$

where

$$c = \frac{B_1}{B_1+B_3} \asymp \frac{\|P_2^T\beta\|^2}{\|P_2^T\beta\|^2+1}. \tag{84}$$

Then theorem follows by the same argument as before.

4.4. Consistency of Test

Since $\beta \sim N(0,\sigma_\beta^2 I_p)$ and $(\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T$ is a projection matrix, we have $\gamma = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\beta \sim N(0,\sigma_\beta^2 I_{n-1})$. Then

$$\begin{aligned}
T &= \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{\gamma^T\tilde{X}^T\tilde{X}\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\gamma^T\gamma + 2\gamma^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\
&= \frac{A_1+A_2+A_3}{B_1+B_2+B_3}.
\end{aligned} \tag{85}$$

It can be seen that $A_1 \sim \|\gamma\|^2\sum_{i=1}^p \lambda_i\chi_1^2 \asymp \|\gamma\|^2(p+\lambda_1) \asymp \sigma_\beta^2 n(p+\lambda_1)$, $A_2 = O_P(\sqrt{A_1})$ and $A_3 \asymp n$. As for the denominator of T, we have $B_1 \asymp \sigma_\beta^2 n$, $B_3 \leq \tilde{\epsilon}(\tilde{X}^T P_2 P_2^T\tilde{X})^{-1}\tilde{\epsilon} \asymp n/p$ and $B_2 = O_P(\sqrt{B_3}\sigma_\beta)$.

By similar technique as before, we have $A_1/B_1 \sim \sum_{i=1}^p \lambda_i\chi_1^2$. By CLT and Slutsky's theorem, we have

$$\frac{\sum_{i=r+1}^p \lambda_i\chi_1^2 - p\lambda}{\lambda\sqrt{2p}} \xrightarrow{\mathcal{L}} N(0,1). \tag{86}$$

And note that if $p\sigma_\beta^2 \to \infty$,

$$
\begin{aligned}
&\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{A_1+A_2+A_3}{B_1+B_2+B_3} - \frac{A_1}{B_1}\Big| \\
&= \frac{1}{\lambda\sqrt{2p}}\Big|\frac{(A_2+A_3)B_1-(B_2+B_3)A_1}{(B_1+B_2+B_3)B_1}\Big| \\
&= \frac{O_P(1)}{\lambda\sqrt{2p}}\Big|\frac{(O_P(\sigma_\beta\sqrt{n(p+\lambda_1)})+O_P(n))O_P(\sigma_\beta^2 n)-(O_P(\sigma_\beta\frac{\sqrt{n}}{\sqrt{p}})+O_P(\frac{n}{p}))O_P(\sigma_\beta^2 n(p+\lambda_1))}{\sigma_\beta^4 n^2}\Big| \\
&= O_P(\frac{p+\lambda_1}{\sigma_\beta\sqrt{np}}) + O_P(\frac{p+\lambda_1}{\sigma_\beta^2 p^{3/2}}).
\end{aligned}
\tag{87}
$$

Hence if

$$
\frac{np^2+p^{5/2}+\lambda_1 p^{3/2}}{(p+\lambda_1)^2}\sigma_\beta^2 \to \infty, \tag{88}
$$

then

$$
\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| = o_P(1). \tag{89}
$$

Equivalently, there exists a positive sequence $\epsilon_n \to 0$ such that

$$
\Pr\Big(\Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| > \epsilon_n\Big) < \epsilon_n. \tag{90}
$$

Then it follows by Lemma 5 and Slutsky's theorem that

$$
\begin{aligned}
&\Pr\Big(\frac{T-p\lambda}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) \\
=&\Pr\Big(\frac{T-p\lambda}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha), \Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \leq \epsilon_n\Big) + o(1) \\
\geq&\Pr\Big(\frac{A_1/B_1-p\lambda}{\lambda\sqrt{2p}} - \epsilon_n \geq \Phi^{-1}(1-\alpha), \Big|\frac{T-\lambda p}{\lambda\sqrt{2p}} - \frac{A_1/B_1-\lambda p}{\lambda\sqrt{2p}}\Big| \leq \epsilon_n\Big) + o(1) \\
=&\Pr\Big(\frac{A_1/B_1-p\lambda}{\lambda\sqrt{2p}} - \epsilon_n \geq \Phi^{-1}(1-\alpha)\Big) + o(1) \\
=&\mathrm{E}\Pr\Big(\frac{\sum_{i=r+1}^p \lambda_i\chi_1^2 - p\lambda}{\lambda\sqrt{2p}} - \epsilon_n \geq \Phi^{-1}(1-\alpha) - \frac{\sum_{i=1}^r \lambda_i\chi_i^2}{\lambda\sqrt{2p}}\Big|\sum_{i=1}^r \lambda_i\chi_i^2\Big) + o(1) \\
=&\mathrm{E}\Phi\Big(-\Phi^{-1}(1-\alpha) + \frac{\sum_{i=1}^r \lambda_i\chi_i^2}{\lambda\sqrt{2p}}\Big) + o(1).
\end{aligned}
\tag{91}
$$

Similarly we get the lower bound. Then

$$\Pr\Big(\frac{T-p\lambda}{\lambda\sqrt{2p}} \geq \Phi^{-1}(1-\alpha)\Big) = \quad \mathrm{E}\Phi\Big(-\Phi^{-1}(1-\alpha)+\frac{\sum_{i=1}^{r}\lambda_i\chi_i^2}{\lambda\sqrt{2p}}\Big)+o(1). \tag{92}$$

4.5. $\hat{\lambda}$

Note that $\tilde{X}^T\tilde{X} = \tilde{X}^T P_1 P_1^T \tilde{X} + \tilde{X}^T P_2 P_2^T \tilde{X}$. By Lemma 2, we have

$$\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X}) \leq \lambda_i(\tilde{X}^T\tilde{X}) \leq \lambda_{r+1}(\tilde{X}^T P_1 P_1^T \tilde{X}) + \lambda_{i-r}(\tilde{X}^T P_2 P_2^T \tilde{X}), \tag{93}$$

for $i \geq r+1$. Note that $\lambda_{r+1}(\tilde{X}^T P_1 P_1^T \tilde{X}) = 0$ since the rank of $\tilde{X}^T P_1 P_1^T \tilde{X}$ is r. Sum the above inequality from $i = k+1$ to $n-1$ and we obtain

$$\sum_{i=k+1}^{n-1}\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X}) \leq \sum_{i=k+1}^{n-1}\lambda_i(\tilde{X}^T\tilde{X}) \leq \sum_{i=k-r+1}^{n-r-1}\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X}). \tag{94}$$

Hence by Lemma 3,

$$\Big|\sum_{i=k+1}^{n-1}\lambda_i(\tilde{X}^T\tilde{X}) - \sum_{i=1}^{n-1}\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X})\Big| \leq k\lambda_1(\tilde{X}^T P_2 P_2^T \tilde{X}) = O_P(p). \tag{95}$$

Since

$$\sum_{i=1}^{n-1}\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X}) = \mathrm{tr}\tilde{X}^T P_2 P_2^T \tilde{X} \sim \lambda\chi^2_{(p-k)(n-1)}, \tag{96}$$

we have by CLT that

$$\sum_{i=1}^{n-1}\lambda_i(\tilde{X}^T P_2 P_2^T \tilde{X}) = \lambda(p-k)(n-1)\big(1+O_P\big(\frac{1}{\sqrt{(p-r)(n-1)}}\big)\big). \tag{97}$$

It follows from (95) and (97) that

$$\frac{1}{(p-r)(n-1)}\sum_{i=k+1}^{n-1}\lambda_i(\tilde{X}^T\tilde{X}) = \lambda + O_P(\frac{1}{\sqrt{np}}) + O_P(\frac{1}{n}) = \lambda + O_P(\frac{1}{n}). \tag{98}$$

When λ is substituted by $\hat{\lambda}$, the conclusion of Theorem 1, 2 and 3 will be still valid if we can prove

$$\Big|\frac{T-\hat{\lambda}p}{\hat{\lambda}\sqrt{2p}} - \frac{T-\lambda p}{\lambda\sqrt{2p}}\Big| \xrightarrow{P} 0. \tag{99}$$

In fact
$$\Big|\frac{T-\hat{\lambda}p}{\hat{\lambda}\sqrt{2p}}-\frac{T-\lambda p}{\lambda\sqrt{2p}}\Big|=\frac{T}{\sqrt{2p}}\frac{|\hat{\lambda}-\lambda|}{\hat{\lambda}\lambda}=O_P(\frac{T}{n\sqrt{p}}). \tag{100}$$

In Theorem 1 and 2, $T = O_P(p)$. Combined with $p = o(n^2)$, it follows that (100) $\xrightarrow{P} 0$.

In Theorem 3, $T = O_P(p+\lambda_1)$. To make (100) $\xrightarrow{P} 0$, we require $(p+\lambda_1)/(n\sqrt{p}) \to 0$.

5. Simulation Results

References

[1] T. T. Cai, Z. Ma, Y. Wu, Sparse pca: Optimal rates and adaptive estimation, Annals of Statistics 41 (6) (2012) 3074–3110.

[2] J. Zhao, X. Xu, A generalized likelihood ratio test for normal mean when p is greater than n, Computational Statistics & Data Analysis.

[3] S. C. Ahn, A. R. Horenstein, Eigenvalue ratio test for the number of factors, Econometrica 81 (3) (2013) 1203–1227.
URL `http://dx.doi.org/10.3982/ECTA8968`

[4] Z. D. Bai, Y. Q. Yin, Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, Annals of Probability 21 (3) (1993) 1275–1294.

[5] R. Durrett, Probability : theory and examples, Journal of the American Statistical Association 87 (418) (2010) 586.

[6] M. L. Eaton, Multivariate statistics: A vector space approach 80 (392) (1983) 72.